AI trustworthiness in prostate cancer imaging: a look at algorithmic and system transparency*

Sara Colantonio, Andrea Berti, Rossana Buongiorno, Giulio Del Corso, Eva Pachetti, Maria Antonietta Pascali, Charalampos Kalantzopoulos, Varvara Kalokyri, Haridimos Kondylakis, Nikolaos Tachos, Dimitris Fotiadis *Senior Member, IEEE*, Valentina Giannini, Simone Mazzetti, Daniele Regge, Nickolas Papanikolaou, Konstantinos Marias, Manolis Tsiknakis *Senior Member, IEEE*

***Abstract*—A responsible approach to artificial intelligence and machine learning technologies, grounded in sound scientific foundations, technical robustness, rigorous testing and validation, risk-based continuous monitoring and alignment with human values is imperative to guarantee their favorable impact and prevent any adverse effects they may have on individuals and communities. An essential aspect of responsible development is transparency, which constitutes a fundamental principle of the European approach towards artificial intelligence. Transparency can be achieved at different levels, such as data origin and use, system development, operation and usage. In this paper, we present the techniques implemented and delivered in the EU H2020 ProCAncer-I project to meet the transparency requirements at the different levels required.**

***Clinical Relevance*—This paper examines the primary transparency hurdles in artificial intelligence for medical imaging diagnostics, and presents the approaches that the EU H2020 project ProCAncer-I is taking to address them.**

I. Introduction

Realizing the potential and advantages of Artificial Intelligence (AI) and Machine Learning (ML) solutions in high-stakes domains, such as medical imaging diagnostics, necessitates high-quality scientific foundations, technical robustness, rigorous testing and validation, risk-based continuous monitoring and responsible development aligned with ethical, legal and social human values. The European approach to AI [1][2] centers around this vision, thus promoting excellence, trust and risk-based management as the primary drivers of a positive impact of AI. In clinical practices, applications can only have a significant impact and uptake if they can ensure reliability and clinical value, gain stakeholders' trust and acceptance, and guarantee the complete safety of patients [3]. One crucial aspect in this respect is transparency. Transparency entails fully documenting the entire lifecycle of an AI system, along with the underlying principles governing its operation. Guaranteeing transparency by design of an AI system is crucial to prevent any ambiguity in how it operates and is utilized by clinical decision makers. As such, it represents a fundamental principle of the FUTURE-AI guidelines [4], specifically addressing Traceability, Explainability, and Usability. Moreover, transparency guarantees that the AI system is reproducible and auditable by design, thus establishing the foundations for accountability and responsibility. In the scientific community, the pursuit of transparency in AI adheres to recently-established tenets such as those of open data and open science. However, in the private sector, it is still a delicate matter, largely due to the competition within the industry. In this brief communication, we present the fundamental aspects of transparency in medical imaging and then provide an overview of the solutions being delivered in the EU H2020 ProCAncer-I project.

II. Transparency: where we stand

AI solutions for medical imaging in oncology heavily rely on data-driven methods using large amounts of multimodal data. In the instance of prostate cancer, multiparametric magnetic resonance examinations are valuable to detect and confirm the presence of a tumor and to extract meaningful information about tumor phenotype [5]. Nonetheless, it is necessary to better integrate them with clinical data (e.g. hormone levels, demographics and medical history) to gather a complete picture of the patient and his risk. Additionally, imaging data could be derived from diverse clinical centers and may differ in terms of clinical standards, acquisition protocols and devices used. For instance, there may be variations in PIRADS versions among different centers. This variation has shown an influence on AI tool's performance, comparable to population attributes [6]. Additionally, various instruments and frameworks can be used to develop inductive AI models, as well as different optimization strategies.

In this multifaceted scenario, transparency must be addressed using competent practices and technical safeguards throughout the AI value chain, from data collection to system development, deployment, and usage.

The principal dimensions of transparency can be categorized as follows:

- *Data transparency*: related to documenting clearly and thoroughly data provenance, including ownership,

*Research partially funded by the ProCancer-I European Union's H2020 programme under Grant Agreement No. 952159

S.C., A.B., R.B., G.D.C, E.P., and M.A.P. are with the Institute of Information Science and Technologies of the National Research Council of Italy, Via G. Moruzzi, 1, 56124 Pisa, Italy (S.C. corresponding author phone: +39 0506213141;e-mail: sara.colantonio@isti.cnr.it).

V.K., H.K., K.M., and M.T. are with the Foundation for Research and Technology Hellas, Institute of Computer Science, Heraklion, Greece.

C.K., N.T. and D.F. are with the Foundation for Research and Technology Hellas, Ioannina, Greece, D.F. is also with University of Ioannina

N.P. is with the Champalimaud Foundation, Computational Clinical Imaging Group, Lisboa, Portugal

V.G., S.M and D.R. are with the Department of Surgical Sciences, University of Turin, 10126 Turin, Italy and with the Department of Radiology, Candiolo Cancer Institute, FPO-IRCCS, 10060 Candiolo, Italy

acquisition methods, reference clinical standards, curation, storage, and processing.

- *System transparency,* related to the need for a complete and standardized report on how the AI system has been developed, validated and tested choices.
- *Decision or algorithmic transparency*, related to the provision of effective clarifications about the logic and the reasoning behind AI classification or prediction results, to enable clinicians to take meaningful actions on it.

Whilst data provenance has been granted standards to follow, model transparency still lacks maturity, despite the partial inclusion of the required functionalities in MLOps frameworks. Decision transparency has gained momentum under the umbrella of eXplainable AI (XAI), which constitutes a renewed effort in decision-making, despite its old roots [7], [8]. XAI is essential for achieving profitable collaboration between humans and AI. However, effective implementation necessitates the use of techniques from various fields (e.g., human-computer interactions). Standards for AI explanations should be sound and aligned with user needs in order to empower them and grant full control over the system.

III. Transparency Measures in ProCAncer-I

In ProCAncer-I, we are taking care of transparency by addressing all its three dimensions.

A. Data and system transparency

In the project, medical and clinical data have been *FAIRified* (i.e., made Findable, Accessible, Interoperable and Reusable) via a GDPR-compliant project infrastructure, which is based on the MOLGENIS meta-data platform. Such a platform functions as the primary metadata catalogue, then allowing users to search clinical and imaging metadata and retrieve cohorts using a combination of different variables. Furthermore, in line with the directives of the FUTURE-AI principles, system transparency is ensured by equipping every AI model with its designated *Model Passport*, housed within a dedicated model registry. The passport presents all the model's information in a uniform format, as a metadata structure. More precisely, the AI Model Passport has been defined as a *minimal provenance model schema* containing information that can be mapped on the following phases of the AI chain:

1. the data collection process, by tracking dataset characteristics and data localization
2. the data processing pipeline, by describing the data transformations that occur from the collected data towards "curated" and harmonized data
3. the model training and validation process, for information about any features extracted from data and their parameters, the training parameters, the evaluation metrics, the testing conditions and the general AI/ML model characteristics that will allow users to monitor as well as enable reproducibility by AI consumers
4. the operation and monitoring of models, by storing performance metrics, uncertainties estimation and metrics used to detect performance drifts.

The current version of the passport is already integrated with renowned tools for data governance and management, such as DVC, as well as MLOps frameworks, as MLflow (see Fig. 1). This way, the metadata fields are automatically populated with the data coming from such tools, without requiring tedious human interventions.

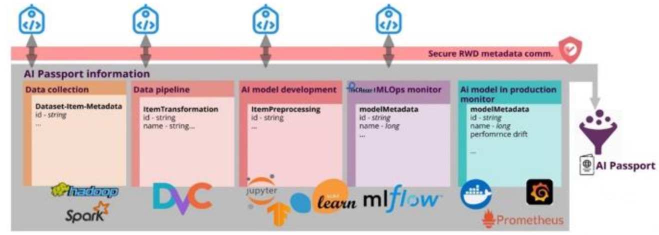

Figure 1. The AI Model Passport with its main components and integration

B. Decision or algorithmic transparency

XAI research seeks to provide human decision makers with explanations to understand the reasoning behind an AI system and its decision-making process. Nonetheless, data-driven algorithm functions according to an implicit problem specification derived from the learning process, which is not immediately apparent nor easily comprehensible by humans. In ProCAncer-I, we run co-design workshops to understand preferences and expectations by clinical end users, with particular focus on the explanation modalities they prefer (see Fig. 2).

Figure 2. Reply statistics for one of the questions asked to radiologists about the format of explanations for the detection use- case

IV. Conclusion

This brief communication introduces the main dimensions of AI transparency and the strategies the EU H2020 project ProCAncer-I is adopting to address them. The authors would like to thank the ProCAncer-I Consortium.

References

[1] EC's Comm "Artificial Intelligence for Europe", COM(2018)237.
[2] EC's Comm "Artificial Intelligence Act", P9_TA(2023)0236
[3] A. Berti, R. Buongiorno, G. Carloni, C. Caudai, G. del Corso, et al. "Exploring the potentials and challenges of Artificial Intelligence in supporting clinical diagnostics and remote assistance for the health and well-being of individuals", Ital-IA, 2023, CEUR Vol. 3486, 2023
[4] K. Lekadir, et al. "FUTURE-AI: Guiding Principles", arXiv preprint arXiv:2109.09658, 2021
[5] Baydoun, A., Jia, A.Y., Zaorsky, N.G. et al. "Artificial intelligence applications in prostate cancer". Prostate Cancer Prostatic Dis, 2023
[6] K Drukker, W. Chen, J.W. Gichoya, N.P. et al. "Toward fairness in artificial intelligence for medical image analysis: identification and mitigation of potential biases in the roadmap from data collection to model deployment. Journal of Medical Imaging, 10(6):061104, 2023
[7] F. Chiarugi, S Colantonio, D Emmanouilidou, M Martinelli, D Moroni and O. Salvetti. "Biomedical signal and image processing for decision support in heart failure". Art Intel in Med, pp. 38-51, Elsevier, 2008
[8] S. Colantonio, *An approach to decision support in heart failure*. In Proc. of SWAP 2007.